Journal of Biological Sciences 14 (4): 267-275, 2014
ISSN 1727-3048 / DOI: 10.3923/jbs.2014.267.275


# Nucleotide Diversity and Association Genetics of Xyloglucan Endotransglycosylase/hydrolase (XTH) and Cellulose Synthase (CesA) Genes in *Neolamarckia cadamba*

[1]S.Y. Tiong, [1]W.S. Ho, [2]S.L. Pang and [1]J. Ismail
[1]Forest Genomics and Informatics Laboratory (fGiL), Department of Molecular Biology,
Faculty of Resource Science and Technology, Universiti Malaysia Sarawak,
94300, Kota Samarahan, Sarawak, Malaysia
[2]Applied Forest Science and Industry Development (AFSID),
Sarawak Forestry Corporation, 93250, Kuching, Sarawak, Malaysia

**Abstract:** A detailed study was carried out to discover single nucleotide polymorphisms (SNPs) from *Neolamarckia cadamba* partial XTH (~1283bp) and CesA (778bp) DNA sequences and further associates those SNPs with basic wood density. Primers were designed in flanking the partial XTH and CesA genes from 15 *N. cadamba* trees. The amplified DNA fragments were sequenced and the basic wood density measurements were determined for each tree. The sequence variation analyses revealed that 34 SNPs (2.65% occurrence) and 3 SNPs (0.39% occurrence) were found in 15 partial genomic DNA sequences of NcXTH1 and NcCesA1, respectively. All the SNPs were discovered in both exon and intron regions. NcXTH1 examined sites showed higher nucleotide diversities of $\pi = 0.00402$ and $\theta_w = 8.919$ when compared to NcCesA1 ($\pi = 0.00127$; $\theta_w = 0.9226$). The LD decayed slowly with distance of polymorphic sites in a linear pattern with the mean $R^2$ value of 0.000687. Association genetics study showed that 2 SNPs from NcXTH1 genes were significantly associated with basic wood density ($p<0.05$) of *N. cadamba*. Once the gene-associated SNP markers in NcXTH1 genes are validated, it could be potentially used as a tool in Gene-Assisted Selection (GAS) of *N. cadamba* trees. This study has also demonstrated that the candidate-gene based association genetics is a powerful approach to dissect complex adaptive traits for organism lacking a genome sequence or reference genomic resources.

**Key words:** *Neolamarckia cadamba*, xyloglucan endotransglycosylase/hydrolase (XTH), cellulose synthase (CesA), single nucleotide polymorphism (SNP), association genetics

## INTRODUCTION

*Neolamarckia cadamba* or locally known as Kelampayan is one of the evergreen and fast growing tropical forest tree species under Rubiaceae family (Joker, 2000). This species is naturally distributed throughout India, Pakistan, Sri Lanka, Burma, Malaysia, Cambodia, Thailand and Laos. The wood is used to make plywood, pulp, paper, boxes, furniture and more. Moreover, *N. cadamba* also served as medicinal plant for traditional curing using its leaf, bark and roof (Patel and Kumar, 2008; Mondal *et al*., 2009; Bussa and Pinnapareddy, 2010). Despite it has high economic value, studies on *N. cadamba* at the molecular level are still limited to date. Recently, a Kelampayan tree transcriptome database (NcdbEST) had been established to provide useful genomics information and resources for researchers to deeply explore the genomics basic of the *N. cadamba* (Ho *et al*., 2010).

Molecular study of wood is important to develop the fundamental study of wood biology and genetics in order to produce 'designer wood' or wood of desired characteristics. DNA markers are now widely used in molecular characterization because DNA markers can characterize cultivars, provenances or genotypes precisely and enable the measurement of genetic relationships (Narayanan *et al*., 2007; Lau *et al*., 2009). Association genetics is the approach that identifies statistical association between phenotypic trait variations and allelic polymorphism in targeted genes. The

**Corresponding Author:** Ho Wei Seng, Forest Genomics and Informatics Laboratory (fGiL), Department of Molecular Biology, Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, 94300, Kota Samarahan, Sarawak, Malaysia Tel: 60 82-581000 Fax: 60 82-583160

association study in plants are generally highlights in the utilization of a wide range of commercial and fitness traits mapped with the well-characterized candidate genes. For instance in maize, candidate genes were associated with digestibility (Guillet-Claude *et al*., 2004), maysin synthesis (Szalma *et al*., 2005) and kernel composition and starch production (Wilson *et al*., 2004). In forest tree species, *Eucalyptus* was the first study carried out to find the relationship between microfibril angle and cinnamoyl CoA reductase (CCR) gene using SNP-association approach (Thumma *et al*., 2005). The first association genetic study using multigenes in forest tree species, *Pinus radiata* was carried out using 58 SNPs from 20 wood and drought-related candidate genes (Gonzalez-Martinez *et al*., 2007). Dillon *et al*. (2010) also examined the allelic variation that affects *Pinus radiata* wood properties using 36 cell wall candidate genes. Gene-associated SNPs found in cinnamate 4-hydroxylase (C4H) and cinnamyl alcohol dehydrogenase (CAD) genes from *Acacia mangium* also showed their significant relationship with wood density, specific gravity and cell wall thickness (Tchin *et al*., 2011).

SNP is the most commonly used new generation molecular marker in breeding programme because it is considered as the most abundant resource for genetic variations studies and are distributed throughout the genome (Halushka *et al*., 1999; Kwok *et al*., 1994). According to Wang *et al*. (1998), SNPs occurs once every 500-1000 bp when comparing two chromosomes. Other estimation also includes one SNP occurs every 100-300 bp in any genome (Gupta *et al*., 2001). SNP is believed to have a relatively stable inheritance as compared to other marker systems. SNP with the low mutation rate makes it an excellent marker to study complex genetic traits, such as wood formation and as a tool for understanding genome evolution (Syvanen, 2001). SNP is also preferred to be used because it serves as a direct marker with the sequence information of the exact nature of the allelic variant is provided (Gupta *et al*., 2001). Technological improvements have made SNP and indel as attractive markers for high-throughput marker-assisted breeding, EST-mapping and the integration of genetic and physical maps (Nasu *et al*., 2002; Rafalski, 2002; Gonzalez-Martinez *et al*., 2007; Ibitoye and Akin-Idowu, 2010; Thomson *et al*., 2010).

Wood is made up of secondary xylem tissues and has a chemical complex of cellulose, lignin, hemicelluloses and extractives. The synthesis and development process of cell walls are very important in wood formation. Xyloglucan endotransglycosylase/hydrolase (XTH) and cellulose synthase (CesA) are two key proteins that play essential role in primary and secondary cell walls, respectively. CesA proteins catalyze cellulose polymerization (Somerville, 2006) and are believed to act as a central catalyst in the generation of plant cell wall biomass (Kumar *et al*., 2009). XTH is considered as a key agent to regulate cell wall expansion and is believed to be responsible for the incorporation of newly synthesized xyloglucan (XG) into the wall matrix (Darley *et al*., 2001).

The selection of high quality wood traits at the early stage is important to produce high quality woods at a low cost (starting modal), energy (labor) and land usage to maximize the commercial value. Association genetics is becoming one of the effective approaches to identify the association between variation in genetic and phenotype. Therefore, identification of gene-associated SNPs that correlated with wood properties (basic wood density) may benefits in marker-assisted breeding programme for plantation establishment. Hence, the objectives of this study were to identify SNPs in partial genomic DNA sequence of XTH and CesA genes and to identify the genetic association of XTH and CesA genes with basic wood density of *N. cadamba*.

## MATERIALS AND METHODS

**Plant materials:** A total of 15 kelampayan trees were selected randomly from the natural stands in Kota Samarahan, Sarawak, Malaysia in October 2011. The inner bark issues were collected and total genomic DNA was isolated by using a modified Doyle and Doyle (1990) protocol. The isolated DNA was purified by using the Wizard Genomic DNA Purification Kit (Promega, USA) according to manufacturer's protocol.

**PCR amplifications:** Primer pair that includes the flanking of complete coding sequence in XTH genomic DNA was designed based on the full-length NcXTH1 cDNA (GenBank accession number: JX134619). For CesA gene, primer pair flanking partial genomic DNA that includes HVRII region was designed based on the complete coding sequence of *NcCesA1* (GenBank accession number: JX134621). All the primers were designed using Primer Premier 5.0 software (PREMIER Biosoft International, USA) as shown in Table 1. PCR reactions were carried out in a total 25 μL reaction volume containing 1x PCR buffer, 0.2 mM dNTPs mix (Invitrogen, Brazil), 1.5 mM $MgCl_2$, 10 pmol of forward and reverse primers each, 1.0 U *Taq* DNA polymerase (Invitrogen, Brazil) and 30 ng of DNA template. The thermal cycling profile was programmed at 94°C for 2 min, 35 cycles at 94°C for 30 sec, 57°C (XTH amplification) or 64°C (CesA amplification) for 45 sec, 72°C for 30 seconds and finally extension at 72°C for 7 min.

Table 1: Primer design for full-length and partial genomic DNA sequences of XTH and CesA, respectively

| Targeted gene | Primer sequence | Amplicon size |
|---|---|---|
| XTH | Forward: 5'- ACAATGGCTTCTCATTTGAACT-3' | 892 bp |
| | Reverse: 5'-TTTGGCTCCTCTCAGATCG-3' | |
| CesA | Forward: 5'-GGTGTTCTTACAAATGCTCC-3' | 1,375 bp |
| | Reverse: 5'- CCATACCAGACAGGGCTA-3' | |

**Cloning and sequencing:** Amplicons were purified using QIAquick® Gel Extraction Kit (QIAGEN, Germany) and ligated into pGEM®-T Easy Vector System (Promega, USA). Clones with targeted inserts were identified by carrying out colony PCR using M13 universal primer pair. Plasmids of positive clones were isolated by using the Wizard Plus SV Minipreps DNA Purification System (Promega, USA). Purified plasmids were then sent for sequencing using 3730xl DNA Analyzer (Applied Biosystem, USA).

**Sequence variation analysis:** Base calling and vector sequences were removed using Chromas version 2.33 (Technelysium, AU). Each of the sequence was verified and checked for their homology using BLASTn provided by NCBI. The sequences were aligned using CLC Main Workbench version 5.0 software (CLC bio, Denmark) to identify single nucleotide polymorphisms. All sequence polymorphisms detected then were visually rechecked from chromatograms. Later, the open reading frame sequences for each gene were translated into amino acid sequences by using ExPASy translate tool (http://web.expasy.org/translate/). Protein sequence alignment was done using CLC Main Workbench version 5.0 (CLC bio, Denmark) to determine the synonymous and non-synonymous mutations.

**Basic wood density measurement:** Wood cores of 5 mm diameter and 8 cm long were sampled from the trunk of each selected *N. cadamba* trees using an increment borer at the height of about 1.3-1.4 m. A replicate was taken for each sample. Wood cores were kept in different labeled tubes placed on ice to maintain the constant humidity for measurement of green value. Green volume of each wood core was taken by measuring the water displaced in the wood core. According to Phytagora's theorem, water has a density of 1. Therefore, the measured weight of displaced water is equal to the sample's volume. Oven-dry weight was measured from the same sample by drying it in a well ventilated oven at 103°C until it achieves constant weight (Chave, 2005). Basic wood density was calculated as oven dry weight per green volume.

**Nucleotide diversity and association genetics analysis:** DNA Sequence Polymorphism version 5 (DnaSP v.5.0) software (Librado and Rozas, 2009) was used to estimate the average number of nucleotide diversity per site ($\pi$), $\theta$ values per site, synonymous and non-synonymous nucleotide polymorphisms and Tajima's D test. Statistical test carried out by DnaSP do not include the insertion and deletion site mutations. TASSEL version 3.0 (Bradbury *et al*., 2007) was used to calculate the $R^2$ value to determine the correlation of SNPs. Linkage disequilibrium (LD) graph was plotted using $R^2$ values against pairwise polymorphic nucleotide distances. General Linear Model (GLM) was also tested using 1,000 permutation test for each SNPs with respect to basic wood density. SNPs that scored at p-value lower than 0.05 considered to be significantly associated with its trait.

## RESULTS AND DISCUSSION

**SNP discovery:** A total of 34 SNPs were found in NcXTH1 genes (~1,290 bp) which corresponds to one SNP in every 38 bp (2.65% occurrence). Similar levels of SNP frequencies were found in other studies. Tchin *et al*. (2012) showed that one SNP occurs in every 59 bp in C4H gene in *N. cadamba*. Krutovsky and Neale (2005) also found that one SNP in every 46 bp in 18 Douglas fir wood quality candidate genes. Another study that used 35 cell wall compound genes recorded one SNP in every 54 bp (Kelleher *et al*., 2012). The large difference of the SNP occurrence might due to the differences in species and number of samples used. For instance, in a study by Dillon *et al*. (2010), one SNP was recorded in every 466 bp of XTH gene with only five samples of *Pinus radiate*.

Out of the 34 polymorphic sites, 20 substitution SNPs and 3 indels were located in the coding regions (exons) and another 9 substitution SNPs as well as 2 indels were located in the non-coding regions (introns and untranslated regions). In this study, more sequence polymorphism was found in the coding sequence (856 bp) because the numbers of sites studied were longer than the non-coding region (434 bp). However, the SNP frequency in non-coding region was slightly higher (39 bp/SNP) than in coding region (37 bp/SNP). Overall, 24 out of 34 (71%) of the SNPs in NcXTH1 were singletons (locus with only one sequence showing a mutated nucleotide) which most of them were located in exons (Table 2). There were only 5 parsimony informative SNPs (locus with two or more sequences owned a nucleotide variant) found in the full-length DNA sequence of NcXTH1.

Full-length NcXTH1 genomic DNA was found to have an equal number of synonymous and non-synonymous SNPs with the proportion of 0.060 each. Although most of the non-synonymous SNPs were found in the coding region, the occurrence of non-synonymous SNPs was highest in non-coding region (1.84%). Nucleotide diversity in NcXTH1 was parallel with the findings of Krutovsky and Neale (2005), where most of the SNPs were singletons, synonymous and found mostly in the non-coding region in 18 wood quality-related candidate genes in Douglas fir trees. Most of the SNPs were caused by transitional mutation (71%), especially the A-G substitution (Table 2). A study by Tchin *et al.* (2012) also shows a higher frequency in transition mutations (55%) in another two wood formation candidate genes, cinnamate 4-hydroxylase (C4H) and cinnamyl alcohol dehydrogenase (CAD) in *N. cadamba*. Tchin *et al.* (2011) also found that A-G transition mutations were frequently found (83%) in partial DNA sequences of CAD in *A. mangium* superbulk trees.

Frequency of SNP in partial DNA sequences of NcCesA1 gene was lower than in NcXTH1 with one SNP detected in every 259 bp. SNP occurrence of CesA gene in *N. cadamba* was predicted to be lower than SNP found in CesA1 of *Shorea parvifolia*, where one SNP was detected in every 115 bp of DNA sequences with only the used of five samples (Seng *et al.*, 2011). On the other hand, NcCesA1 showed higher SNP frequency than in CesA genes of *Pinus radiate*. Dillon *et al.* (2010) reported that one SNP was detected in every 403 and 374 bp DNA sequences of CesA1 and CesA3, respectively, in *Pinus radiate*. In this study, very low SNP frequency was detected in the coding region, with only 1 SNP detected in every 531 bp.

Table 3 shows the summary of SNP distribution in partial NcCesA1 DNA sequences (778 bp). A total of three SNPs found in NcCesA1 were caused by substitution mutation and no Indel was detected. Two out of three SNPs located in non-coding regions. In reverse with NcXTH1, more transversion SNPs found in NcCesA1 with one G-C and G-T substitution detected each. All of the three SNPs found in NcCesA1 partial DNA sequences were synonymous, with higher frequency in the non-coding region (0.80%). Meanwhile, only one singleton SNP was found in the exon of NcCesA1 while two parsimony informative SNPs found in the non-coding regions.

**Nucleotide diversity analysis:** Sequence statistical analysis of NcXTH1 and NcCesA1 using DnaSP v.5.0 software (Librado and Rozas, 2009) was examined on all the sequences (sites) excluding gaps and missing data. All of the calculation and analysis did not include Indels. Alignment of 15 NcXTH1 full-length DNA sequences contained 13 sites with alignment gaps and thus, 1,277 out of 1,290 sites of NcXTH1 full-length DNA sequences were examined. In NcCesA1 partial DNA sequences alignment, no gap was found and all of the 778 sites were included in the analysis. Nucleotide diversity, Tajima's test and minimum number of recombination events of NcXTH1 and NcCesA1 are summarized in Table 4.

Table 2: Distribution of SNPs in NcXTH1 gene

| Characteristic | Coding region | Non-coding region | Total |
|---|---|---|---|
| No. of sites examined (bp) | 856 | 434 | 1,290 |
| No. of polymorphic sites | 23 | 11 | 34 |
| Average SNP frequency (bp/SNP) | 37 | 39 | 38 |
| Singleton SNP | 18 | 6 | 24 |
| Parsimony informative SNP | 2 | 3 | 5 |
| Transition mutation | 17 | 7 | 24 |
| Transversion mutation | 2 | 3 | 5 |
| Deletion mutation | 3 | 2 | 5 |
| Synonymous SNP | 14 | 3 | 17 |
| Non-synonymous SNP | 9 | 8 | 17 |
| Frequency of synonymous SNP (%) | 1.64 | 0.69 | 1.32 |
| Frequency of non-synonymous SNP (%) | 0.01 | 1.84 | 1.32 |

Table 3: Distribution of SNPs in NcCesA1 gene

| Characteristic | Coding region | Non-coding region | Total |
|---|---|---|---|
| No. of sites examined (bp) | 529 | 249 | 778 |
| No. of polymorphic sites | 1 | 2 | 3 |
| Average SNP frequency (bp/SNP) | 529 | 83 | 259 |
| Singleton SNP | 1 | 0 | 1 |
| Parsimony informative SNP | 0 | 2 | 2 |
| Transition mutation | 1 | 0 | 1 |
| Transversion mutation | 0 | 2 | 2 |
| Synonymous SNP | 1 | 2 | 3 |
| Frequency of synonymous SNP (%) | 0.19 | 0.80 | 0.39 |

Table 4: Summary statistics of sequence variations and Tajima's D test for both NcXTH1 and NcCesA1

| Characteristics | NcXTH1 | NcCesA1 |
|---|---|---|
| Total No. of sites (bp) | 1,290 | 778 |
| No. of sites examined (bp) | 1,277 | 778 |
| No. of polymorphic (segregating) sites, S | 29 | 3 |
| Total No. of mutation, Eta (n) | 30 | 3 |
| Nucleotide diversity, $\pi$ | 0.00402 | 0.00127 |
| $\theta$ (per site) from $\pi$ | 0.00404 | 0.00128 |
| $\theta$(per site) from S | 0.00708 | 0.00119 |
| $\theta$ (per site) from n | 0.00723 | 0.00119 |
| Average No. of nucleotide differences, k | 5.133 | 0.9905 |
| $\theta$ (per sequence) from S, $\theta_w$ | 8.919 | 0.9226 |
| Tajima's D | -1.867* | 0.2187 |
| Tajima's at silent sites | -1.667 | 0.21873 |
| Tajima's D at coding region | -2.008* | -1.159 |
| Tajima's D of synonymous SNP (Syn) | -1.911* | -1.159 |
| Tajima's D of non-synonymous SNP (NonSyn) | -1.864* | - |
| Tajima's D (NonSyn/Syn) ratio | 0.9753 | - |

InDel were excluded for all measurement; * indicates the significant value ($p<0.05$)

Nucleotide diversity ($\pi$) is the average number of nucleotide differences per site between two sequences and the population mutation parameter is represented by $\theta$. Nucleotide diversity of 15 DNA sequences of NcXTH1 was estimated to be higher than NcCesA1 at 0.00402 and 0.00127, respectively. Overall, all of the $\theta$ values per site calculated from number of nucleotide diversity ($\pi$), total number of mutation (n), or polymorphic sites (S) in NcXTH1 were much more higher than in NcCesA1 (Table 4). The $\theta$ value per sequence estimated from S($\theta_w$) of NcXTH1 (8.919) was almost 10-fold higher than in NcCesA1 (0.9226). This was caused by the higher frequency of SNP in NcXTH1 than in NcCesA1. $\theta_w$ value of both genes examined in *N. cadamba* were predicted to be much more higher than other wood-related candidate genes that scored $\theta_w$ lower than 0.02 (Brown *et al*., 2004; Krutovsky and Neale, 2005).

Tajima's D statistic reflects the difference between nucleotide diversity, $\pi$ and theta ($\theta$) per sequence from number of polymorphic sites (S), $\theta_w$. The value of D is expected close to zero at equilibrium between genetic drift and selectively neutral mutation (Brown *et al*., 2004). In NcXTH1, almost all of the SNPs showed significant Tajima's D value (less than 0.05), except those at silent (synonymous and non-coding) sites. The negative values (-1.667 to -2.008) indicate an excess of low frequency variants are segregating in NcXTH1 gene. Similarly, most of the Tajima's D values were negative in wood quality-related candidate genes in *Pseudotsuga menziesii* (Krutovsky and Neale, 2005) and in *Pinus taeda* (Brown *et al*., 2004). In contrast to NcXTH1, all of the Tajima's D values tested in NcCesA1 failed to show significant results with an overall positive value of 0.2187.

**Linkage disequilibrium (LD):** A considerable amount of Linkage Disequilibrium (LD) was found within NcXTH1 and NcCesA1 sequences using TASSEL software (Bradbury *et al*., 2007). A total of 496 pairwise comparisons were estimated from the examined sites. More than 80% of the paired sites showed LD statistically significant by a two-sided Fisher's Exact test. $R^2$ values were calculated which represents the correlation between alleles at two loci and is informative for evaluating the resolution of association approaches (Bradbury *et al*., 2007). In this study, the mean $R^2$ value for 496 estimates of LD was 0.000687. Average $R^2$ value of *N. cadamba* was far lower than in another timber species balsam poplar (0.52; Olson *et al*., 2010) and in soybean (0.36; Zhu *et al*., 2003).

Averaged across all loci, appreciable decay of LD to the values <0.10 within 350 bp was detected when $R^2$ was plotted against distance between sites in base pairs for NcXTH1 and NcCesA1 sequenced gene fragments (Fig. 1). As compared to other species, *N. cadamba* do not show significant decay when only two genes were used in LD estimation. Nucleotide studies on other species showed that LD decayed more than 50% including loblolly pine (Brown *et al*., 2004), soybean (Zhu *et al*., 2003) and Douglas fir tree (Krutovsky and Neale, 2005). According to a study by Nordborg (2000), LD decays more rapidly in outcrossing species as compared to selfing species because recombination is more effective in outcrossing species.

**Basic wood density:** The diameter at breast height (dbh) of *N. cadamba* trees was in the range of 24.0 cm to 102.0 cm with an average of 39.2 cm (Table 5). Most of the *N. cadamba* trees are having the diameter between 24.0 to 50.0 cm, except for sample NcMT1 (102.0 cm). The green weight, oven dry weight and green volume of each replicate were taken and recorded to calculate the mean basic wood density. On average, the basic wood density of *N. cadamba* was 368.93 kg $m^{-3}$. Wood sample NcMT1 recorded the highest basic density (444.03 kg $m^{-3}$) while

Table 5: Diameter (dbh) and basic wood density measures of 15 N. cadamba trees

| Sample | dbh (cm) | Basic wood density (kg $m^{-3}$) |
|---|---|---|
| NcMT1 | 102.0 | 444.03 |
| NcMT2 | 34.5 | 348.78 |
| NcMT3 | 28.5 | 383.68 |
| NcHT1 | 59.7 | 391.10 |
| NcHT2 | 42.5 | 346.73 |
| NcHT4 | 24.0 | 300.25 |
| NcHT5 | 27.1 | 315.43 |
| NcGT1 | 25.8 | 339.70 |
| NcAT1 | 32.3 | 395.90 |
| NcAT3 | 37.4 | 377.46 |
| NcKT1 | 41.0 | 360.98 |
| NcST1 | 28.9 | 381.11 |
| NcST2 | 39.6 | 343.68 |
| NcUT1 | 40.1 | 397.33 |
| NcLT1 | 25.0 | 407.83 |

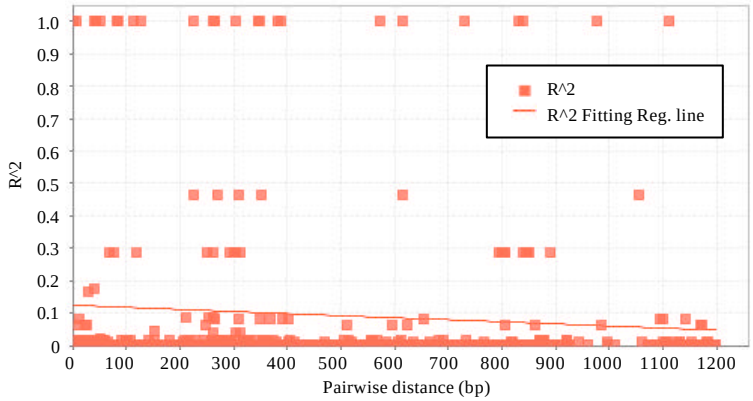


Fig. 1: Plot of the squared correlations of allele frequencies ($R^2$) versus nucleotide distance in base pair between polymorohic sites (SNPs) in NcXTH1 and NcCesA1. The straight line represents $R^2$ fitted regression line

NcHT4 recorded the lowest (300.25 kg $m^{-3}$) among 15 samples. Basic density was determined because it is one of the most important criteria in breeding programmes to evaluate wood potential for pulp, paper or wood industries (Seca and Domingues, 2006).

**Association genetics study:** Out of 34 SNPs detected in NcXTH1, two of them (SNP8 and SNP29) showed significant association ($p<0.05$) with the basic wood density at position 118 and 1,173 bp, respectively (Fig. 2). Both markers were located in exons and caused by synonymous T-C transition mutation. None of the non-synonymous SNP was found to be significantly associated with the basic density. The associated-SNP8 and-SNP29 were found to occur in the individual tree NcMT1 that had the highest wood density (444.03 kg $m^{-3}$). A similar study also had been carried out on the association of SNP markers in another two wood candidate genes (C4H and CAD) with the basic wood density (Tchin *et al*., 2011, 2012). In this study, two SNPs were also found to be significantly associated with the basic wood density of *N. cadamba* and *A. mangium*. This supports the findings that SNP8 and SNP29 could potentially be developed into SNP markers for superior *N. cadamba* tree selection once validated in the future.

Previous studies have shown that there is a strong relationship between wood quality-related candidate genes with the wood properties (Plomion *et al*., 2000; Wegrzyn *et al*., 2010; Tchin *et al*., 2011, 2012). Plomion *et al*. (2000) had shown that the growth of maritime pines was affected by the biochemical content, such as lignin and cellulose content, where these chemical contents were controlled by its genetics. Association genetics of traits controlling lignin and cellulose biosynthesis in *P. trichocarpa* also have been demonstrated by the rapid decay of within-genes (SuSy1 and C4H1) LD and the high coverage of amplicons across each gene (Wegrzyn *et al*., 2010). These data reflected that numerous identified polymorphisms are in close proximity to the causative SNPs.

In conclusion, this study has demonstrated that association genetics is a powerful tool to discover

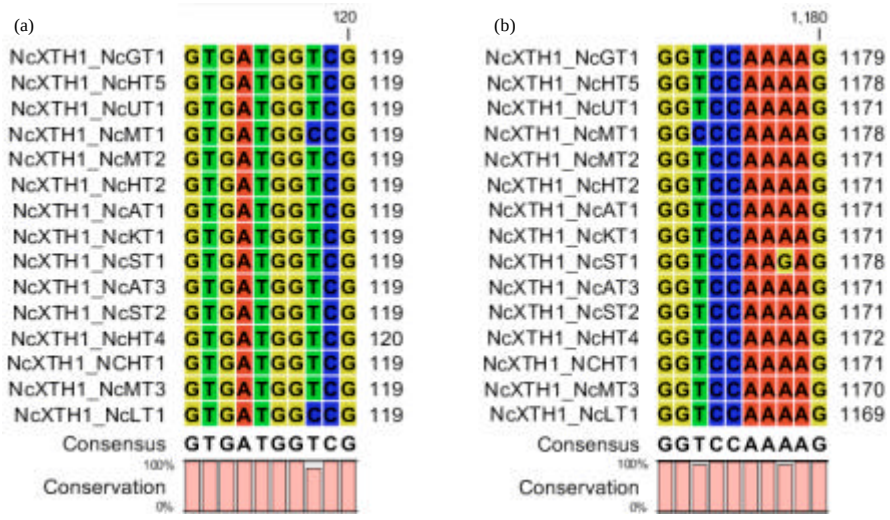


Fig. 2(a-b): Position of significant basic wood density associated SNPs at position (a) 118 bp and (b) 1,173 bp

naturally occurring allelic variations in genes associated with economically important traits by harnessing the genetic variation at the population level. The gene-associated SNP identified in XTH gene of *N. cadamba* could be potentially used as a tool in Gene-Assisted Selection (GAS) of *N. cadamba* after validated in breeding programmes to produce quality planting materials that are of faster growth, high-yield and high wood quality and also adapted to local conditions, so that we may achieve economic benefits of great significance.

## ACKNOWLEDGMENTS

The authors would like to thank all the lab assistants and foresters involved in this research project for their excellent field assistance in sample collection. This study is part of the joint Industry-University Partnership Programme, a research programme funded by the Sarawak Forestry Corporation and Universiti Malaysia Sarawak (Grant No. 02(DPI09)832/2012(1)), RACE/a(2)884/20121(02) and GL(F07)/06/2013/STA-UNIMAS(06)).

## REFERENCES

Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss and E.S. Buckler, 2007. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics, 23: 2633-2635.

Brown, G.R., G.P. Gill, R.J. Kuntz, C.H. Langley and D.B. Neale, 2004. Nucleotide diversity and linkage disequilibrium in loblolly pine. Proc. Natl. Acad. Sci. USA., 101: 15255-15260.

Bussa, S. K. and J. Pinnapareddy, 2010. Antidiabetic activity of stem bark of *Neolamarckia cadamba* in alloxan induced diabetic rats. Int. J. Pharm. Technol., 2: 314-324.

Chave, J., 2005. Measuring wood density for tropical forest trees: A field manual for the CTFS sites. Lab. Evolution et Diversite Biologique, University Paul Sabatier, France, February 16, 2005, pp: 1-7.

Darley, C.P., A.M. Forrester and S.J. McQueen-Mason, 2001. The molecular basis of plant cell wall extension. Plant Molecular Biol., 47: 179-195.

Dillon, S. K., M. Nolan, W. Li, C. Bell, H. X. Wu and S. G. Southerton, 2010. Allelic variation in cell wall candidate genes affecting solid wood properties in natural populations and land races of *Pinus radiate*. Genetics, 185: 1477-1387.

Doyle, J.J. and J.L. Doyle, 1990. Isolation of plant DNA from fresh tissue. Focus, 12: 13-15.

Gonzalez-Martinez, S.C., N.C. Wheeler, E. Ersoz, C.D. Nelson and D.B. Neale, 2007. Association genetics in *pinus taeda* L. I. wood property traits. Genetics, 175: 399-409.

Guillet-Claude, C., C. Birolleau-Touchard, D. Manicacci, P. M. Rogowsky and J. Rigau *et al*., 2004. Nucleotide diversity of the ZmPox3 maize peroxidase gene: Relationships between a MITE insertion in exon 2 and the variation in forage maize digestibility. BMC Genetics, 10.1186/1471-2156-5-19

Gupta, P.K., J.K. Roy and M. Prasad, 2001. Single nucleotide polymorphisms: A new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. Curr. Sci., 80: 524-535.

Halushka, M.K., J.B. Fan, K. Bentley, L. Hsie and N. Shen *et al*., 1999. Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. Nat. Genet., 22: 239-247.

Ho, W.S., S.L. Pang, P.S. Lai, S.Y. Tiong and S.L. Phui *et al*., 2010. Genomics studies on plantation tree species in Sarawak. Proceedings of the International Symposium on Forestry and Forest Products 2010: Addressing the Global Concerns and Changing Societal Needs, October 5-7, 2010, Kuala Lumpur, Malaysia, pp: 172-182.

Ibitoye, D.O. and P.E. Akin-Idowu, 2010. Marker-assisted-selection (MAS): A fast track to increase genetic gain in horticultural crop breeding. Afr. J. Biotechnol., 9: 8889-8895.

Joker, D., 2000. *Neolamarckia cadamba* (Roxb.) Bosser (*Anthocephalus chinensis* (Lam.) A. Rich. ex Walp.). Seed Leaflet No. 17, September 2000, Danida Forest Seed Centre, Denmark, pp: 1-2.

Kelleher, C.T., J. Wilkin, J. Zhuang, A.J. Cortes and A.L.P. Quintero *et al*., 2012. SNP discovery, gene diversity and linkage disequilibrium in wild populations of *Populus tremuloides*. Tree Genet. Genomes, 8: 821-829.

Krutovsky, K.V. and D.B. Neale, 2005. Nucleotide diversity and linkage disequilibrium in cold-hardiness and wood quality-related candidate genes in Douglas fir. Genetics, 171: 2029-2041.

Kumar, M., S. Thammannagowda, V. Bulone, V. Chiang and K.H. Han *et al*., 2009. An update on the nomenclature for the cellulose synthase genes in *Populus*. Trends Plant Sci., 14: 248-254.

Kwok, P.Y., C. Carlson, T.D. Yager, W. Ankener and D.A. Nickerson, 1994. Comparative analysis of human DNA variations by fluorescence-based sequencing of PCR products. Genomics, 23: 138-144.

Lau, E.T., W.S. Ho and A. Julaihi, 2009. Molecular cloning of *cellulose synthase* gene, SpCesA1 from developing xylem of *Shorea parvifolia* spp. *parvifolia*. Biotechnology, 8: 416-424.

Librado, P. and J. Rozas, 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics, 25: 1451-1452.

Mondal, S., G.K. Dash and S. Acharyya, 2009. Analgesic, anti-inflammatory and antipyretic studies of *Neolamarckia cadamba* barks. J. Pharm. Res., 2: 1133-1136.

Narayanan, C., S.A. Wali, R. Shukla, R. Kumar, A.K. Mandal and S.A. Ansari, 2007. RAPD and ISSR markers for molecular charazterization of teak (*Tectona grandis*) plus trees. J. Trop.Sci., 19: 218-225.

Nasu, S., J. Suzuki, R. Ohta, K. Hasegawa and R. Yuia *et al*., 2002. Search for and analysis if single nucleotide polymorphisms (SNPs) in rice (*Oryza sativa*, *Oryza rufipogon*) and establishment of SNP markers. DNA Res., 9: 163-171.

Nordborg, M., 2000. Linkage disequilibrium, gene trees and selfing: An ancestral recombination graph with partial self-fertilization. Genetics, 154: 923-929.

Olson, M.S., A.L. Robertson, N.Takebayashi, S. Silim, W.R. Schroeder and P. Tiffin, 2010. Nucleotide diversity and linkage disequilibrium in balsam poplar (*Populus balsamifera*). New Phytol., 186: 526-536.

Patel, D. and V. Kumar, 2008. Pharmacognostical studies of *Neolamarckia cadamba* (Roxb.) Bosser leaf. Int. J. Green Pharm., 2: 26-27.

Plomion, C., C. Pionneau, J. Brach, P. Costa and H. Bailleres, 2000. Compression wood-responsive proteins in developing xylem of maritime pine (*Pinus pinaster* Ait.). Plant Physiol., 123: 959-969.

Rafalski, A., 2002. Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol., 5: 94-100.

Seca, A.M.L. and F.M.J. Domingues, 2006. Basic density and pulp yield relationship with some chemical parameters in *eucalyptus* trees. Pesquisa Agropecuaria Bras., 41: 1687-1691.

Seng, H.W., P.S. Ling, P. Lau and I. Jusoh, 2011. Sequence variation in the cellulose synthase (SpCesA1) gene from *Shorea parvifolia* ssp. *parvifolia* mother trees. Pertanika J. Trop. Agric. Sci., 34: 317-323.

Somerville, C., 2006. Cellulose synthesis in higher plants. Annu. Rev. Cell Dev. Biol., 22: 53-78.

Syvanen, A.C., 2001. Accessing genetic variation: Genotyping single nucleotide polymorphisms. Nat. Rev. Genet., 2: 930-942.

Szalma, S. J., E.S. Buckler, M.E. Snook and M.D. McMullen, 2005. Association analysis of candidate genes for maysin and chlorogenic acid accumulation in maize silks. Theor. Appl. Genet., 110: 1324-1333.

Tchin, B.L., W.S. Ho, S.L. Pang and J. Ismail, 2012. Association genetics of the Cinnamyl alcohol dehydrogenase (CAD) and Cinnamate 4-hydrolase (G4H) genes with basic wood density in *Neolamarckia cadamba*. Biotechnology, 11: 307-317.

Tchin, B.L., W.S. Ho, S.L. Pang and J. Ismail, 2011. Gene-associated single nucleotide polymorphism (SNP) in cinnamate 4-hydroxylase (C4H) and Cinnamyl alcohol dehydrogenase (CAD) genes from *Acacia mangium* superbulk trees. Biotechnology, 10: 303-315.

Thomson, M.J., K. Zhao, M. Wright, K.L. McNally, H. Leung and S.R. McCouch, 2010. Development and application of 96- and 384-plex Single Nucleotide Polymorphism (SNP) marker sets for diversity analysis, mapping and marker-assisted selection in rice. Proceedings of the 2nd Africa Rice Congress: Innovation and Partnerships to Realize Africa's Rice Potential, March 22-26, 2010 Bamako, Mali, pp: 1-5.

Thumma, B., R., M.F. Nolan, R. Evans and G.F. Moran, 2005. Polymorphisms in Cinnamoyl CoA Reductase (CCR) are associated with variation in microfibril angle in *Eucalyptus* sp. Genetics, 171: 1257-1265.

Wang, D.G., J.B. Fan, C.J. Siao, A. Berno and P. Young *et al*., 1998. Large-scale identification, mapping and genotyping of single-nucleotide polymorphisms in the human genome. Science, 280: 1077-1082.

Wegrzyn, J.L., A.J. Eckert, M. Choi, J.M. Lee and B.J. Stanton *et al*., 2010. Association genetics of traits controlling lignin and cellulose biosynthesis in black cottonwood (*Populus trichocarpa*, Salicaceae) secondary xylem. New Phytol., 188: 515-532.

Wilson, L.M., S.R. Whitt, A.M. Ibanez, T.R. Rocheford, M.M. Goodman and E.S. Buckler, 2004. Dissection of maize kernel composition and starch production by candidate gene association. Plant Cell, 16: 2719-2733.

Zhu, Y.L., Q.J. Song, D.L. Hyten, C.P. Van Tassell and L.K. Matukumalli *et al*., 2003. Single-nucleotide polymorphisms in soybean. Genetics, 163: 1123-1134.